DOCUMENT RESUME

ED 169 128 TM 008 665

AUTHOR Peck, Robert F.; And Others
TITLE Student Evaluation of Teaching: A Multivariate Validation Study.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Aug 78
CONTRACT NE-C-00-3-0066
GRANT OB-NIE-G-78-0116
NOTE 11p.; Paper presented at the Annual Meeting of the American Psychological Association (86th, Toronto, Ontario, Canada, August 28-31, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Anglo Americans; Black Students; *Effective Teaching; Elementary Education; *Elementary School Teachers; Grade 6; Mexican Americans; *Student Evaluation of Teacher Performance; *Teacher Behavior; Test Bias; Test Reliability; *Test Validity
IDENTIFIERS Student Evaluation of Teaching

ABSTRACT

Student Evaluation of Teaching (SET), a six-scale questionnaire, yielded an overall teaching effectiveness score for 53 teachers, rated by 1,128 Anglo, black, and Chicano students in their sixth grade social studies classes in 1974-75; and for 42 teachers (33 of them the same) in 1975-76 reading classes. SET correlated substantially with analogous attitude measures; was free of socio-economic status or ethnic bias; and correlated .38 to .68 with observers' teacher ratings. The SET scores for the first year correlated .68 with the SET scores for the second year. Teacher personality and behavior affected change in SET, as expected, and SET positively affected student outcomes. Almost all findings were replicated in the second year. (Author/MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: TM SP
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED169128

STUDENT EVALUATION OF TEACHING: A MULTIVARIATE VALIDATION STUDY

Robert F. Peck, Abraham Blattstein, Ronald Fox

Research and Development Center for Teacher Education

The University of Texas at Austin

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PRE-PUBLICATION COPY-NOT FOR DISTRIBUTION

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R+D Center for Teacher Education

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

TM008 665

This study was supported in part by National Institute of Education Contract NE-C-00-3-0066, Grant Number OB-NIE-G-78-0116, Research and Development Center for Teacher Education. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education and no official endorsement by that office should be inferred.

Paper to be presented at the annual meeting of the American Psychological Association to be held in Toronto, Canada, August 1978.

STUDENT EVALUATION OF TEACHING:

A MULTIVARIATE VALIDATION STUDY

Student Evaluation of Teaching (SET), a six-scale questionnaire, yielded an overall "teaching effectiveness" score for 53 teachers, rated by 1128 students in their randomized, tri-ethnic sixth grade social studies classes in 1974-75; and for 42 teachers (33 of them the same) in 1975-76 reading classes. SET correlated substantially with analogous attitude measures, were free of SES or ethnic bias; correlated .38 to .68 with observers' teacher ratings (Ryans, 1975). The SET scores for year one (1974-75) correlated .68 with the SET scores for year two (1975-76). Teacher personality and behavior affected change in SET, as expected, and SET positively affected student outcomes. Almost all findings replicated.

STUDENT EVALUATION OF TEACHING:
A MULTIVARIATE VALIDATION STUDY

Are student descriptions of teacher effectiveness valid? Students, in the typical school, are the only people who regularly see a teacher at work, every day. While all kinds of human bias could be expected to affect an individual student's judgement, pooling the observations of a whole class should tend to reduce subjective distortion, and approach a true picture of how the teacher acts, and what impact this has on students. Practically speaking, an economical measure of this kind, if it were valid, could provide a teacher with feedback for purposes of self-evaluation and self-development. Such feedback from a pool of observers is not practiced in most schools; nor is it economically feasible to provide multiple adult observers to all teachers, as a tool for in-service education.

With such purposes in mind, Veldman and Peck began in the early 1960's to evolve the Student Evaluation of Teaching (SET) instrument. In various forms, it has proved possible to measure several different aspects of teaching, and also to derive a stable, overall estimate of "teacher effectiveness." In one study of fifty teachers who were independently rated by two classes of junior high school students, correlations were computed across the paired classes for five SET factor scores. The coefficients ranged from .72 to .92 (Veldman, 1970); a considerable indication that the scores might be validly describing these teachers.

In the current Teaching-Learning Interaction Study (Peck, 1977), a six-scale Student Evaluation of Teaching (SET) instrument was used that yielded one strong "effectiveness" score. The average ratings by all students in a class on a scale were summed across the six scales to derive the SET score for each teacher. This was done about six weeks into the school year, in

October (pretest), and again in late April (posttest). Internal consistency (alpha) was .73 for the pretest, and .81 for the posttest (pupil N=1128). Test-retest reliability across two weeks was .78 (pupil N=61). This process was carried out with the Social Studies classes of 53 sixth grade teachers in Austin, Texas in 1974-75, and with the reading classes of 42 teachers in 1975-76. Thirty-three of the teachers participated in both years, and were thus evaluated by two successive groups of students, on different subject matter classes. (Another 27 teachers in Daviess County, Kentucky, took part in the first year, 22 of whom also took part in the second year. Their data remain to be analyzed, at this date.) The Austin student population was about 67% Anglo, 19% Black, 17% Chicano, and 1% Oriental. The students were randomly assigned to classes, so that each class approximated this distribution closely.

SET was part of a network of instruments designed to study the interacting effects of numerous teacher and student characteristics on a number of outcomes of a year's schooling (achievement, coping skill, attitudes toward school, self-esteem, etc.). Providing feedback to teachers concerning their teaching effectiveness can facilitate self-evaluation and self-development which, in turn, should affect the teaching process. It is therefore logical to assume that the students' evaluation of teaching effectiveness as feedback for purposes of improved teaching would be valid since they are affected by the teaching process. In addition, the students have extended contact with the teacher. The purpose of the study was to provide empirical evidence supporting the proposition that students' evaluation of teaching effectiveness is valid in terms of freedom from bias, and the following types of validity: construct, cross rater, consequent, predictive, and replicative.

Construct Validity was tested by correlating the SET-post score with attitude toward school and teacher as measured by the School Sentiment Index

(SSI) (UCLA Instructional Objectives Exchange). The test-retest reliability of SSI was .87 (N=151, t=2 weeks). In the Austin sample, this reliability was .84 (N=63, t=2 weeks). A further check on construct validity was obtained by correlating SET post with a year two score on How This Class Makes Me Feel, a set of items analogous to many items in the Piers-Harris Self-Concept Questionnaire, but asking specifically how a particular class (teacher) affected the student's feelings and perceptions about himself as a student (test-retest reliability .78, N=60, t=one week.)

Freedom from Bias: Tests were made to see if the ethnicity or the SES of the student raters had a biasing effect on the SET scores they individually gave their teachers at the beginning of the year or on the posttest SET scores, (adjusted for pretest). In both cases, the analysis used a linear regression model. The pretest was used as a covariate whenever the posttest was the criterion. Both linear and quadratic terms were entered for ethnicity and SES, plus terms for the interactions of ethnicity and SES. All student pretest measures were tested for their effects on the SET score.

Cross Rater Validity: Several of the measures were specifically designed to test the validity of SET in several ways. Ryans' COR rating system (Ryans, 1960) was used by classroom observers, during the year, yielding highly reliable (.97, .98, .98) scores for each teacher, across several observations, on the factors Kindly-Understanding (KU), Systematic-Organized (SO), and Stimulating-Imaginative (SI). Correlating these ratings by experienced teachers with SET scores in each year tests the cross rater validity of both sets of judges.

Complete data in both years were available for 28 teachers who were rated by the students in their social studies classes (in most cases) in 1974-75, and by a new group of students, in a reading class, in 1975-76. Differences in

the characteristics of the students, the difference in subject matter, and a year's difference in the teachers made the correlation of SET scores a quite stringent test of the validity of SET as a purported measure of teacher effectiveness as a stable, generalizable pattern.

Consequent Validity: Changes in SET over the year were related to Ryans' three dimensions of teacher behavior (KU, SO, SI) by regression analysis, to see if, as would logically be expected, teachers' classroom behavior effected changes in their evaluation by the students (regressed change on SET). This tests the "consequent" validity of post-SET by seeing if it is affected by preceding teacher behavior.

Regressed change in SET was also used to see whether self-reported teacher personality characteristics, assessed early in the year, had an effect on the evaluation of the teachers by the students. The assessment instruments were the Adjective Self-Description (ASD) measuring seven traits, and the Views of Teaching. The ASD has reported test-retest reliability from .80 to .92, depending on the scale (t=2 weeks) and is reported to be valid (Veldman, 1971). Views of Teaching is a sentence-completion form, adapted from an earlier measure (Peck, 1973), with two scales: Coping Effectiveness and Attitudes. Test-Retest reliability is .73 and .53. respectively.

Predictive Validity: The individual SET scores were regressed on each student outcome, to test the influence of students' individual perceptions of teacher on their change in achievement, attitude, self-esteem and coping skills.

By the time this paper is to be presented, an analysis will also be completed which tests the predictive validity of the class mean SET score. The teachers post-SET score will be regressed on each of the student outcome mea-

sures, to determine whether the teacher SET score affects achievement, attitudes and coping skills changes over the year, as it logically should.

Replicative Validity: The final test was one of replication. Each of the tests described was performed separately with the 1974-75 sample and the 1975-76 sample. Each kind of validity finding was checked to see if it occurred similarly in the two samples, which had many of the same teachers in common, but entirely different sets of children. (They were different, not just in identity, but in many of their behaviors and attitudes.)

RESULTS

Construct Validity: SET-post correlated .67 and .57 with SSI-Post, in years one and two respectively. SET-post and How This Class Makes Me Feel correlated .49 in year two. These similar but not identical attitude measures showed substantially the amount of relationship to be expected if the SET were a valid estimate of teacher impact on the students.

Freedom from Bias: There were no ethnic or SES effects on SET-pre, or on regressed change in SET. Further, SET-pre was not affected by student achievement level (or by students' peer-rated academic coping skills). The students were not subjectively biased by any of these factors in their own background or performance, when they evaluated their teachers.

Cross Rater Validity: SET-post (class mean) correlated with the observers' ratings on COR-KU, COR-SO, and COR-SI, .68, .54, and .47 in year one; .54, .45, and .38 in year two respectively. The students and the adult observers agreed to a considerable degree in describing the teachers, in both years. Taken with the evidence of lack of bias, this suggests that the students are giving a reasonably valid picture of teacher behavior, and its effects on them. This is further supported by the finding that the different classes in

two successive years agreed .68 in rating the same set of 28 teachers.

Consequent Validity: SET was influenced, as it logically should be, both by teacher personality characteristics and by teacher classroom behavior. Six of nine personality self-report scales showed effects on regressed change on SET, $p < .05$, in year one; 3 of 9, in year two. The classroom behavior measures, as would be expected, showed stronger, positive effects on SET-change. The percentages of regressed change in SET explained in year one were COR-KU 5.1%, COR-SO 3.1%, COR-SI 3.7%. In year two, the figures were COR-KU 1.3%, COR-SO .8%, COR-SI 2.4%. All of these, except COR-SO in year two, fell below the $p < .01$ level. Both teacher self-reports and observer ratings tended to confirm the validity of the students' assessments.

Predictive Validity: The SET-pre individual score (not the class means) did influence student regressed change on the several cognitive and non-cognitive outcome measures. SET-pre made a significant contribution ($p < .01$, $\geq$ 1.0% variance) to positive change in attitude toward school (SSI), general attitude toward life (SSC-A), coping skills (SSC-C), and self-esteem (P-H). It did not affect achievement change; but, for reasons too many to discuss here, standardized achievement tests, by their design, appear to be extremely ill-suited to detect teacher influence. The strongest test will be the prediction of student effects by the teacher's (class mean) score on SET.

Replicative Validity: Each of the kinds of positive validity evidence found in the 1974-75 sample were again found, independently, in the 1975-76 sample.

IMPLICATIONS

SET appears to be substantially valid as an evaluative description of teaching behaviors for Anglos, Blacks, and Chicanos alike. These sixth grade students were unprejudiced in their appraisals. Their judgments seem to have been based on realistic observations of their teachers' behavior. At least, they agreed significantly with the judgments of the experienced-teacher observers. Finally, SET change was positively affected by just those teacher traits and observed actions that logically should affect it; and, so far, SET is showing logically expectable effects on student outcomes.

Some limited experimentation with private feedback of such data to teachers suggests that, although it is initially worrisome to some teachers, most take the feedback quite seriously. Even in cases where SET does not give them a happy message about their past teaching, quite a few of them take it as a constructive challenge. Some have notably improved their general style of teaching in a subsequent year. Good scores are a great moral booster, of course. Left undisturbed by feedback, most teachers appear to be pretty consistent from year to year, as measured by both SET and COR (Peck, Olsson, and Green, 1978).

References
Student Evaluation of Teaching: A Multivariate Validation Study

Peck, R.F. Teaching-learning interaction study: progress report (3 volumes). Austin: University of Texas, Research and Development Center for Teacher Education, 1977.

Peck, R.F., Olsson, G., and Green, J.L. The consistency of individual teaching behavior. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Peck, R.F. A replication study of coping patterns in eight countries. Austin: University of Texas Personality Research Center, 1973. (ERIC ED 078 342).

Veldman, D.J. Adjective self-description, RMM-11R. Research methodology monographs No. 11., Austin, Texas: Research and Development Center for Teacher Education, 1971.

Veldman, D.J. Student evaluation of teaching. RMM-10. Research methodology monograph No. 10. Austin, Tx: Research and Development Center for Teacher Education, 1970.